**Guided Capstone Project Report**

**Big Mountain Resort Pricing Model**

**Problem Statement**

What actionable strategies can be developed and implemented to increase Big Mountain Resort (BMR) ticket prices and final revenues by at least $1,540,000 before the next ski season in response to the recent rise in operation costs?

**Motivation and Success Metrics**

BMR is currently analyzing data to enhance its ticket pricing strategy and increase its revenue. The ultimate objective is to identify the best ticket price and evaluate the impacts of the proposed changes. Given the recent increase in operating expenses, this is crucial as the resort needs to optimize its revenues and control costs. Within six months, actionable recommendations will be provided for implementation before the next ski season. The success of this work will be determined by whether the recommended ticket pricing strategy results in maximized revenue without significantly impacting visitor numbers (currently at 350K per year). Additionally, the proposed changes should effectively optimize costs, particularly regarding the operational costs of the new chair lift ($1,540,000). Any data-driven changes that are implemented must maintain or improve current service levels.

**Data Munging, Exploratory Data Analysis (EDA), and Feature Engineering**

BMR plan to use a predictive model to guide pricing decisions and future investments.

Firstly, we imported the data from the resorts around the country (330 rows and 27 columns), including BMR in Montana. By checking the missing values, we realized some of the features have missing values. We made sure there were no duplicates. After checking the “state” and “Region” features, we found out that the data contains 38 Regions and 35 states, and Montana is ranked 13th among states with the highest number of resorts, which raises the question of whether the State should be used as a predictor feature in our ticket price model.

By comparing “AdultWeekday” and “AdultWeekend” ticket prices, we realized that for Montana, the prices are the same during weekdays and weekends, and therefore we dropped the “AdultWeekday” since it had more missing values. We defined “AdultWeekend” as the target. Also, we dropped the “fastEight” column since half the values are missing and the rest are 0. This gave us a clean dataset of 277 rows and 25 columns for “ski\_data” and a dataset of 35 rows and 8 columns for the “state\_summary”.

We investigated the EDA to find out whether there is any pattern between the features. There were 2 datasets, “ski\_data” and “state\_summary”, which had 1 categorical field (“state”) and 7 numerical fields. We realized that although New York has the most resorts in our dataset, it is not in the top 5 for the most skiable area. We dropped 2 fields that only contained state-specific data and added 2 more that represent the ratio of the number of resorts to the given population and given area.

We applied PCA on the “state\_summary” dataset to create new features and simplify the visualization in a lower dimension. It was revealed that the first 2 components account for over 75% of the variance (Figure 1), and since there was not any specific pattern for the ski summaries of each state with ticket price (Figure 2), we decided to treat all the states equally.

The “ski\_data” dataset has 25 columns, including 2 categorical fields (“Name” and “Region”). We merged the 2 datasets, and generated 4 features, namely “resort\_skiable\_area\_ac\_state\_ratio”, “resort\_days\_open\_state\_ratio”, “resort\_terrain\_park\_state\_ratio” and “resort\_night\_skiing\_state\_ratio”, and dropped the same state-related features. The descriptive analysis of the final version of “ski\_data” dataset is presented in Table 1.

A picture containing text, line, plot, screenshot

Description automatically generated

Figure 1. Cumulative variance ratio explained by PCA components.

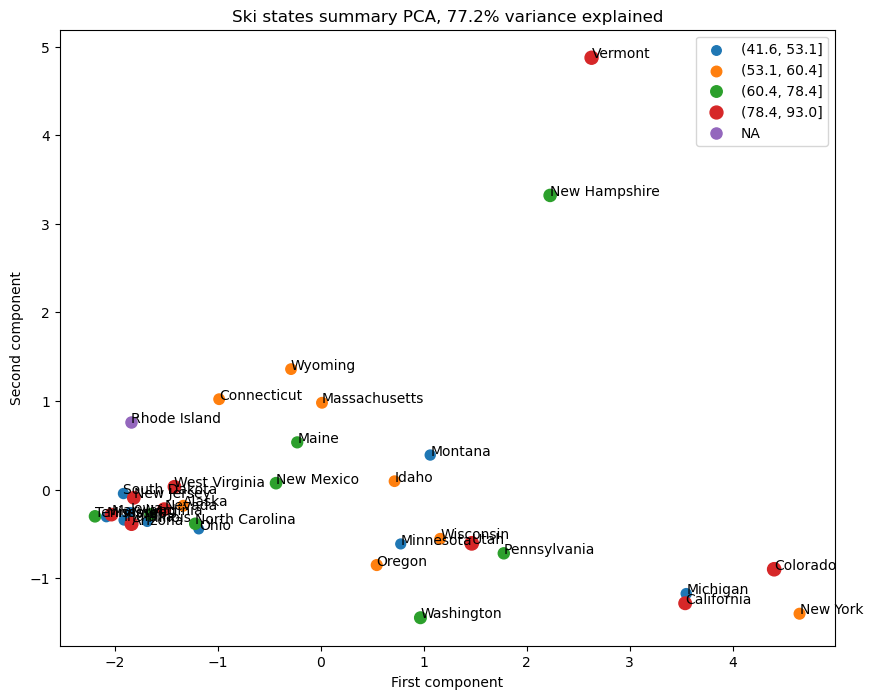


Figure 2. Ski state summary PCA.

**Building the model with evaluation metrics and scenario modeling**

After splitting data into the training set (70%) and the test set (30%), we defined our model as the mean baseline and calculated different metrics. Then we applied “LinearRegression” with “StandardScaler” and the imputation methods using both median and mean values to train the model. We considered R-squared or the coefficient of determination, mean absolute error (MAE), and mean squared error (MSE) as the evaluation metrics. Assessing the model revealed 0.816 and 0.667 of R-squared or coefficient of determination for the training set and test set, respectively.

Table 1. Final descriptive statistics of “ski\_data” dataset.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| **summit\_elev** | 277 | 4310.386 | 3601.05 | 440 | 1400 | 2400 | 7330 | 13487 |
| **vertical\_drop** | 277 | 1106.473 | 843.7064 | 60 | 450 | 825 | 1600 | 4425 |
| **base\_elev** | 277 | 3192.051 | 3064.54 | 100 | 850 | 1360 | 5840 | 10800 |
| **trams** | 277 | 0.111913 | 0.4404 | 0 | 0 | 0 | 0 | 3 |
| **fastSixes** | 277 | 0.093863 | 0.358942 | 0 | 0 | 0 | 0 | 2 |
| **fastQuads** | 277 | 0.722022 | 1.737894 | 0 | 0 | 0 | 1 | 15 |
| **quad** | 277 | 0.945848 | 1.310752 | 0 | 0 | 0 | 2 | 8 |
| **triple** | 277 | 1.440433 | 1.563009 | 0 | 0 | 1 | 2 | 8 |
| **double** | 277 | 1.776173 | 1.835759 | 0 | 0 | 1 | 2 | 14 |
| **surface** | 277 | 2.440433 | 1.744834 | 0 | 1 | 2 | 3 | 9 |
| **total\_chairs** | 277 | 7.530686 | 4.518476 | 0 | 5 | 6 | 9 | 40 |
| **Runs** | 274 | 43.58394 | 37.64129 | 3 | 19 | 32 | 55 | 336 |
| **TerrainParks** | 233 | 2.532189 | 1.618866 | 1 | 1 | 2 | 3 | 10 |
| **LongestRun\_mi** | 272 | 1.333824 | 1.055173 | 0 | 0.5 | 1 | 2 | 6 |
| **SkiableTerrain\_ac** | 275 | 540.1418 | 910.7281 | 8 | 82.5 | 180 | 592.5 | 8464 |
| **Snow Making\_ac** | 240 | 140.0542 | 138.6167 | 2 | 46.5 | 100 | 184.5 | 700 |
| **daysOpenLastYear** | 233 | 111.6738 | 32.71976 | 3 | 95 | 110 | 130 | 243 |
| **yearsOpen** | 277 | 57.90253 | 17.07129 | 6 | 50 | 58 | 70 | 104 |
| **averageSnowfall** | 268 | 175.1716 | 131.4952 | 18 | 65 | 131 | 252.5 | 669 |
| **AdultWeekend** | 277 | 64.27632 | 24.59988 | 17 | 47 | 60 | 78 | 179 |
| **projectedDaysOpen** | 236 | 117.428 | 29.1416 | 30 | 100 | 117.5 | 136 | 233 |
| **NightSkiing\_ac** | 163 | 95.18405 | 100.1034 | 2 | 38.5 | 70 | 104.5 | 650 |

To avoid overfitting, we applied “SelectKBest” with its score function “f\_regression`” along with “cross-validation” and “GridSearchCV”. Results suggested that considering 8 features for the input would give us the best performance. Moreover, training the model using the “RandomForest” model along with the “feature importances” for tuning the hyperparameters revealed the top 4 features (Figure 3), namely “fastQuads”, “Runs”, “Snow Making\_ac”, and “vertical\_drop”. We compared the “LinearRegression” with “RandomForest” to pick the best estimator, which was the “RandomForest” model since it had lower cross-validation MAE by almost $1.

A picture containing text, screenshot, diagram, plot

Description automatically generated

Figure 3. Random forest regressor feature importance.

BMR has been charging $81 for the ticket, but based on the model, they can increase the ticket price to $95.97, and since the calculated mean absolute error (MAE) is $10.50, this suggests there is room for an increase. Understanding the relationship between BMR and ticket prices is essential for business intelligence. Therefore, BMR has narrowed down the possibilities to 4 specific scenarios considering that the expected number of visitors over the season is 350K, and visitors tend to ski for 5 days:

* Permanently closing up to 10 of the least used runs. The model showed that closing one run makes no difference. Closing 2 and 3 successively reduce ticket price support and revenue drastically. If BMR closes 3 runs, they may as well close to 7, as there's no further loss in the ticket price. Increasing the closures down to 8 or more leads to a significant drop (Figure 4).
* Increasing the vertical drop by adding an extra run to the point of 150 ft lower down and installing an additional chair lift without additional snow-making coverage. By this scenario, BMR can increase the ticket price by $1.45, resulting in $2,545,455 more revenue over the season.
* Same as the previous one but adds 2 acres of snow-making cover. The model showed no significant impact on support for increasing ticket prices and revenue compared to scenario 2, indicating that such a small increase in the snow-making area makes no difference.
* Increasing the longest run by 0.2 mi requires an additional snow-making coverage of 4 ac. Based on the model, this change would not make any difference.

A picture containing screenshot, plot, diagram, text

Description automatically generated

A picture containing text, diagram, line, plot

Description automatically generated

Figure 4. Predicted ticket price and the revenue change by the number of runs closed.

**Recommendations**

Based on modeling from other resorts' price strategies, it was determined that the BMR ticket price of $81 is 17.3% lower than the predicted price of $95, and since the MAE is $10.50, we recommend that BMR increase the price.

Moreover, after investigating the proposed scenarios, we suggest increasing the vertical drop by 150 ft, which would support the increase of price by $1.45 and produce $2,545,455 more revenue over the season.

Also, we recommend that BMR provide operation cost data for further analysis and future consideration in modeling.

**Conclusion and Future Work**

Our model and analysis can be used by business leaders for the following:

* Pricing Strategies, to predict ticket prices based on various factors, BMR can devise more effective pricing strategies, optimize revenue, and improve profitability.
* Market Analysis, to simulate and analyze different market scenarios, which can inform strategic planning and forecasting.
* Customer Classifications, to assist in identifying distinct customer groups with different price sensitivities, which can be used for targeted marketing and sales strategies.

To make this model accessible to business analysts, a user-friendly interface, such as a web application or a tool within an existing business intelligence platform, can be used. This will enable analysts to input various feature values and receive predictions without needing knowledge of the underlying machine learning code. Additionally, this interface can allow analysts to adjust parameters, test different scenarios, and view the impact of their changes in real time. This self-service analytics tool will allow analysts to conduct their own exploratory analysis and scenario planning.